

## Method for speech synthesis

Speech processing methods are known, for example, from  
10 US 6 029 135, US 5 732 388, DE 19636739 C1 and  
DE 19719381 C1. Text stored in non-spoken form can be  
output as speech via speech synthesis. As a rule, for  
this purpose a search is made for the individual words  
of the text in a database which contains the phonetic  
15 transcriptions of numerous words. The phonetic  
transcriptions of the words found in the database are  
combined and can be output as speech.

It is also known to assemble the phonetic transcription of a given word from the phonetic transcriptions of its subwords when the given word consists exclusively of these subwords.

Starting from here, it is the object of the invention to improve speech synthesis to the effect that it is



possible to a greater extent to have recourse to phonetic transcriptions of words specified in a database, and that OOV treatments need be used only to a lesser extent.

5

This object is achieved by means of a method, an arrangement and a computer program product having the features of the independent patent claims.

- 10 It is possible by means of the method, the arrangement or the computer program product to have recourse to the phonetic transcriptions of the subwords of a given word even when the given word cannot be assembled completely from subwords contained in the database. The essential
- 15 idea in this case is that use is made for the first time of a hybrid mode of procedure in which both the phonetic transcription of complete subwords, and an OOV treatment are used for the same given word.
- 20 In a preferred development, the OOV treatment for phonetic transcription of the further constituent is performed as a function of the phonetic transcription of the subword found. This renders it possible to markedly raise the quality of the speech synthesis for
- 25 the further constituent by comparison with a corresponding pure OOV treatment of the entire word. The reason for this is firstly that the phonetic transcription of the subword found is very much more reliable than a phonetic transcription of this subword
- 30 by an OOV treatment would be. Consequently, it is possible to proceed from a reliable phonetic context in the OOV treatment of the further constituent, and this permits the OOV treatment to come to the correct result with a very much higher probability. Secondly, the
- 35 phonetic transcription of the subword found is very much longer than the phonemes normally used in an OOV treatment. For this reason, the phonetic context is not only more reliable, but also longer, and so OOV

T01E30" 95/24650



treatment for the further constituent can be carried out on the basis of a larger amount of relevant information. However, this advantage need not necessarily be utilized for the claimed preferred development. Under specific conditions, it can also be sensible when for the OOV treatment for phonetic transcription of the further constituent as a function of the phonetic transcription of the subword found account is taken only of the part of the subword which is immediately adjacent to the further constituent.

The method becomes particularly advantageous when it is not interrupted after a first subword has been found, but a search is made for still further subwords in the given word. This way, as large a section as possible of the given word is assembled from subwords for which reliable information is present in the database, and only the remaining, mostly small further constituent of the word need be subjected to an OOV treatment.

If this remaining further constituent is between two subwords found, the OOV treatment is preferably undertaken as a function of both subregions found. Specifically, in this case both the left-hand and the right-hand phonetic context of the further constituent are reliably prescribed, for which reason it is possible to carry out the OOV treatment with excellent results.

The search for subwords in the database can be optimized by means of various measures. Thus, for example, the aim might be to search only for subwords which have a prescribed minimum length. In practice, a length of 5 letters has proved to be the minimum length, it also being possible for minimum lengths of 3, 4 or 6 letters to be sensible in the case of other boundary conditions, for example for a different language.



Furthermore, the search result is improved when the search for a word part of the given word is not immediately interrupted after the first matching subword is found, but a search is further made for other possible subwords. This can be performed, for example, by supplementing the word part with further letters. As a rule, with this mode of procedure the best result is produced when the longest subword is selected from a plurality of subwords found. However, it is also possible to select a shorter subword when, in conjunction with a longer subword found in the database and contained in the given word, this shorter subword constitutes a larger part of the given word than does the longer subword found per se, when the latter cannot be combined with the second subword found.

The OOV treatment for phonetic transcription of the further constituent can be performed by means of a neuron network.

Alternatively or in addition, a rule-based method or a DTW method can be used for the OOV treatment for phonetic transcription of the further constituent. Such a method is described, for example, in Rüdiger Hoffmann "Signalanalyse und -erkennung" ["Signal analysis and recognition"], Springer Verlag, Berlin, 1998.

However, the OOV treatment can also be performed by means of a second database which contains the phonetic transcription of filling particles normally used in the case of composite words. In German, these are particularly dative and genitive endings which are appended in composite words to the word respectively occurring in front.



Further essential features and advantages of the invention follow from the description of an exemplary embodiment, with the aid of the drawing, in which:

- 5 figure 1 shows a schematic of the cycle of the method,  
and  
figure 2 shows a schematic of a further constituent,  
occurring between two subwords, of a given  
word.

10

The method is to be explained with reference to the example of the given German word "Trainingslager" ["training camp"]. A search is to be made only for subwords with a minimum length of five letters. In step  
15 S1 in accordance with figure 1, a search is made for subwords of the given word in a database which contains phonetic transcriptions of words. Since the minimum length is set to five letters, a start is made by searching for the word "Train". This word is not found  
20 in a German language database. If the database also contains English language words, the first subword of the given word has already now been found. However, a further search is preferably made not only in the first, but also in the second case. This is performed  
25 by searching for the word "Traini". This letter combination is not found in the database. The same holds for the letter combination "Trainin" for which a search is made thereafter.

30 By contrast, the nearest letter combination "Training" is found in the database. Nevertheless, in this case, as well, a further search is preferably made, specifically for the letter combination "Trainings" and the longer letter combinations, formed in the  
35 corresponding continuation of this search step, of the given word. Assuming that the given word "Trainingslager" is not found in its entirety in the

T07E89" 9E24650



database, no further subwords are found in the database.

For the case of an English language and German language database, the longer subword "Training" is selected from the two subwords found, namely "Train" and "Training". This selection step does not occur in the example of a purely German language database.

10 The phonetic transcription registered in the database is selected in step S3 for the subword "Training" found.

It is stipulated in accordance with step S4 that in addition to the subword "Training" found the given word "Trainingslager" has a further constituent "slager" which is not registered in the database.

This further constituent "slager" is then transcribed phonetically in step S5 by means of an OOV treatment. This OOV treatment is preferably based on a conversion of the individual graphemes of the further constituent "slager" into phonemes by means of a neuron network. The phonemes are selected and combined by the neuron network so as to produce the best possible speech synthesis for the further constituent per se.

For an even better speech synthesis result, the OOV treatment for phonetic transcription of the further constituent "slager" is performed as a function of the phonetic description, selected from the database, of the subword "Training" found. In the example selected, the subword "Training" found, or its phonetic transcription reliably prescribes the left-hand phonetic context of the further constituent "slager". The neuron network used for the OOV treatment of the further constituent "slager" can therefore proceed from a reliable result of the syllables of the given word



which preceded the further constituent, and can supply a correspondingly reliable result for the phonetic transcription of the further constituent.

- 5 Finally, in the last step S6 of the method for speech synthesis the phonetic transcription of the subword "Training" found and the phonetic transcription of the further constituent "slager" are combined.
- 10 The speech synthesis result can be further improved when a search is made not only for subwords beginning from the start of the given word, but the search is also started from other areas of the given word. If a specific minimum length  $i$  is prescribed for the
- 15 subword, it is to be recommended to start the further search with the  $i$ +first letter. In the given example, the further search is then started for  $i=5$  with the letter sequence "ingsl" which, for its part, is also of the given minimum length. This letter sequence would
- 20 not be found in the database. The same holds for the letter sequences "ingsla", "ingslag" etc. for which a search is made thereafter.

- Since no subword of any sort is found during this
- 25 further search, the search following thereupon is started not with the letter  $2*i+1$ , but already with  $i+2$ . However, the search sequence "ngsla", "ngslag" etc. also leads to no result. After further corresponding searches have been carried out, however,
- 30 the further subword "lager" is found in the last search. This further subword "lager" found does not originate from the word part of the word "Trainingslager" for which the first subword "Training" was found. Consequently, there is no need in the
- 35 example to select between the two subwords.

Rather, it is now the letter "s" which remains as further constituent of the given word "Trainingslager".



This single letter "s" can be phonetically transcribed very easily by means of an OOV treatment. In this case, there is a further alleviating circumstance that in accordance with figure 2 both the left-hand context 1 "Training" and the right-hand context 3 "lager" are known for the center 2 "s".

Instead of the OOV treatment by means of a neuron network, as was described above, it is also possible in this case for the OOV treatment to be performed by a search in a further database in which the phonetic transcriptions of filling particles normally used with composite words are contained. The genitive s of the present example is such a filling particle normally used. It would therefore be found in the second database, and the associated phonetic transcription would be selected.

Alternatively, however, it is also possible to use rule-based methods and DTW methods for the OOV treatment. In each case, better phonetic transcriptions of the further constituent are to be expected when the phonetic transcription of a plurality of or all subwords found is taken into account in the OOV treatment for phonetic transcription of the further constituent. Of course, this is the case, in particular, when the further constituent in the word is arranged between two subwords found.

Finally, in a last step the phonetic transcription of the subword "Training" found, the phonetic transcription of the further subword "lager" found and the phonetic transcription of the further constituent "s" are then combined for speech synthesis.

35

The arrangement according to the invention can be implemented in the form of a computer system which is programmed to execute a corresponding method.